Cross-linguistic transfer in academic journal writing: Preliminary evidence from lexical bundle analysis in Russian and English

Maria Kostromitina
Northern Arizona University / USA

Abstract – Lexical bundles are considered a fundamental feature of academic writing and have been extensively studied by corpus linguists. However, while learner corpus-based studies have noted the differences between first (L1) and second languages (L2) in the production of lexical bundles, few of them have assessed the underlying causes of such differences, particularly regarding cross-linguistic transfer. The present study investigates the use of lexical bundles in professional writing in the field of Educational Psychology produced by L1 English and L1 Russian authors in order to evaluate the evidence of cross-linguistic transfer in the writing of L2 English learners with L1 Russian background and examine the patterns of L2 English lexical bundle use that mirror L1 English production. This exploratory study compares the frequency and discourse functions of lexical bundles produced by native speakers of English to those used by Russian speakers in their L2 English professional writing, as well as professional writing in their L1. The results of the study indicate that L2 English writing produced by Russian speakers displays overlap in the composition and use of lexical bundles in L1 Russian writing pointing at possible L1 transfer.

Keywords – lexical bundles; professional writing; L1 transfer; cross-linguistic analysis; Russian; English

1. INTRODUCTION

Linguistic features of academic writing registers have been scrutinized by corpus researchers over the last few decades (see Hyland 2015). One of those features are lexical bundles (LBs), or recurrent lexical sequences identified through corpus analysis (Pan *et al.* 2016). As described by Paquot (2013), lexical bundles may be grammatically complete or incomplete phrasal (e.g., *at the same time*, *the results of the*) or clausal (e.g., *I think that*, *is used as the*) segments that fulfil certain discourse functions. As such, LBs have been found to generally act as referential markers (e.g., *at the end of*), text organizers (e.g., *as shown in figure*), and stance markers (e.g., *it is possible that*) in written registers (Biber *et al.* 2003). Corpus linguistic studies have often compared native speaker (L1) and second language (L2) learner production of LBs in academic writing in a target

language (e.g., Chen and Baker 2010; O'Donnell *et al.* 2013). One major limitation of such studies, however, is that they do not consider the native language of the writers and its possible influences on the way lexical bundles are patterned in academic texts. In other words, the common approach contrasting L1 and L2 LBs in a target language without examining the third component, academic writing in learners' L1, does not provide us with insights into the possible reasons behind the observed LB patterns. While studies have recognized the active role that cross-linguistic transfer may play in L2 writing (e.g., Bybee 2008; Paquot 2014), evidence of transfer has been limited. Moreover, the existing evidence of cross-linguistic transfer has been largely inconsistent in motivating the approaches to the assessment of transfer in previous studies and thus potentially weakening the validity of the results.

The goal of the present study is to explore the potential L1 influence in L2 English professional writing produced by Russian authors through the analysis and comparison of high-frequency LBs in three corpora of academic articles in the field of Educational Psychology published in L1 English, L2 English, and L1 Russian. The purpose is to contribute to our understanding of the use of LBs in L2 published academic writing and provide insights into the possible causes of discrepancies in the use of LBs in L1 and L2 writing. First, the study compares the patterns of LB use in L1 and L2 English writing produced by Russian writers to provide further evidence regarding the development of L2 academic writing. Second, the role of L1 influence in the use of LBs by L2 English learners is examined. More specifically, the study compares the use of LBs in two varieties of writing within one discipline: L2 English written by Russian native speakers and L1 Russian. Applying Jarvis's (2000) intra-L1-group congruity criterion of the L1 influence identification framework, the study aims to determine the extent to which the use of LBs in L2 English academic writing made by Russian native speakers differs from the L1 Russian norms. To that end, a functional analysis of LBs in the three language varieties (L1 English, L2 English, L1 Russian) is also performed to collect additional evidence of L1 influence in published L2 English writing.

To the author's knowledge, this is the first study that examines evidence of L1 English-likeness and possible L1 transfer in the use of LBs by Russian speakers of English. Importantly, the study examines expert writing from the discipline of Educational Psychology to avoid confounding "register/discipline differences with the

difference between groups of writers" (Pan *et al*. 2016: 62). More precisely, the study addresses the following research questions:

1. To what extent does L1 Russian writers' use of high-frequency LBs in L2 English writing in the field of Educational Psychology can be attributed to writers' L1 influence and/or proximity to L1 English production?

   a. What are the bundles that are shared between the three language varieties?

   b. What are the bundles that are shared between L1 and L2 English writing?

   c. What are the bundles that are shared between L2 English and L1 Russian writing? (Jarvis's (2000) Intra-L1-group congruity criterion).

2. What are the differences in discourse functions of the identified LBs used in L1 English, L2 English, and L1 Russian expert writing in Educational Psychology?

The paper starts with a brief overview of current literature on LBs in academic writing as well as the role of writers' L1 in LB use (Section 2). Section 3 follows with a description of the corpus used in the study as well as the methodology used. Finally, the results are presented and discussed in Section 4 with regard to Jarvis's (2000) framework of L1 transfer as well as within the domain of academic writing.

2. OVERVIEW OF THE LITERATURE

2.1. LBs in academic writing

Since the introduction of the concept of LBs, or "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status" (Biber *et al*. 1999: 990), studies in corpus linguistics have examined their role in L2 writing (e.g., Simpson-Vlach and Ellis 2010; Salazar 2014). Particularly in the domain of English for Academic Purposes (EAP), researchers have agreed that learners' control of formulaic sequences, such as LBs, is essential for successful academic writing as this register exhibits "a distinct set of lexical bundles, associated with [its] typical communicative purposes" (Biber and Barbieri 2007: 265). Further investigating this argument, Hyland (2008) explored the forms, structure, and functions of LBs in a large corpus of academic writing within four disciplines. He found that bundles were not only important for academic discourse, but also for differentiating texts by discipline (Hyland 2008: 57). Increasingly, in the field of EAP, studies have used this framework to compare and analyze the use of LBs by native

speakers and L2 learners of English in academic writing. So far, it has been shown that the use of formulaic language largely depends on the language level of L2 writers. For example, Staples *et al*. (2013) examined learners' use of bundles in prompted Test of English as a Foreign Language (TOEFL) writing tasks. The study showed that high proficiency learners used fewer bundles compared to low proficiency learners, thus lending support to the hypothesis that learners move towards self-constructed rather than formulaic language with an increase in their target language proficiency (Ellis 2002).

Although LBs are likely to be observed in advanced academic writing, it is still unclear whether highly proficient L2 English learners use them effectively. Research seems to agree that learners misuse L1 English bundles and fail to understand their pragmatic functions in agreement with L1 conventions (Granger 1998; Nekrasova 2009). For example, Chen and Baker (2010) compared LBs retrieved from a corpus of published academic texts with LBs in two corpora of student academic writing (L1 and L2). The study demonstrated that L2 learners employed a smaller range of LBs in their writing; furthermore, they overused certain expressions which were rarely used by native speakers (Chen and Baker 2010: 43). Adapting Chen and Baker's (2010) methodology, Ädel and Erman (2012) investigated the use of English-language LBs in advanced learner writing in comparison with native-speaker writing. For their analysis, the researchers focused on writing by undergraduate university students in the discipline of linguistics. The study found that native speakers included a larger and more varied number of LB types in their writings, including negations, unattended *this*-bundles, existential *there*-bundles, and hedging bundles (Ädel and Erman 2012: 86).

Regarding the discourse functions of LBs in L2 writing, English learners' language production has been found to exhibit lack of register awareness, as well as phraseological and semantic misuse (Gilquin *et al*. 2007; Paquot 2014). Pan *et al*. (2016) conducted a corpus-driven analysis of LBs used by L1 English and L2 English (L1 Chinese) academic professionals writing for telecommunications research journals. The study found major structural and functional differences in LBs between L1 and L2 writing. More specifically, L1 and L2 professionals employed structurally different bundles serving similar functions (Pan *et al.* 2016: 69). On the other hand, a few studies have argued that the use of LBs in L1 and L2 academic writing is largely similar (Swales and Feak 2004; Wulff and Römer 2009). Claims have been made that even though L2 English writers

overuse high frequency LBs, they use the same amount of bundles as L1s overall (Durrant and Schmitt 2009).

In sum, many learner corpus-based studies have noted the differences in L1 and L2 production and use of LBs in discourse. Emphasizing the frequency information of L1 and L2 bundles, studies have explained patterns of overuse or underuse of LBs in learner texts (e.g., Gilquin 2008; Chen and Baker 2010). However, research has largely overlooked the possible underlying explanations for learner deviations in LB use as well as approaches to the investigation of these explanations. In other words, although the findings of the studies mentioned above are valuable in that they provide insights into the differences in the use of LBs in L1 and L2 writing, they do not necessarily investigate the possible causes behind the observed discrepancies. The following section provides an overview of current research of one of such causes, namely, L1 influence.

2.2. L1 influence in the use of LBs

It has been hypothesized that misuse of LBs in an L2 is in part related to L1 influence or transfer, defined as a statistically significant process "occurring from the native language to the foreign language" (Jarvis 2000; see also Selinker 1966: 103; Odlin 2003). One way of investigating such an influence in L2 writing has been Contrastive Interlanguage Analysis. The aim of such analysis is to identify the over- and under-use of chosen features (i.e., LBs) in L2 learners' production in order to detect L1 interference (Granger 2002; Rica Peromingo 2012). For instance, Lu and Deng (2019) compared the use of LBs in dissertation abstracts written by doctoral students who were L1 English speakers and L2 English learners from China. The four-word bundles identified in the study were categorized structurally and functionally revealing substantial differences in the frequencies of use across categories. More specifically, Chinese students demonstrated an underuse of bundles containing indefinite articles that the authors linked to the lack of the article system in Chinese. In a similar study, Esfandiari and Barbary (2017) contrasted four-, five-, and six-word LBs in psychology research articles written by L1 English and L2 English speakers from Iran. The study found that Persian writers used fewer LBs overall and in structurally and functionally different ways when compared to L1 English writers. As such, Persian writers utilized significantly more dependent clauses and significantly fewer research-oriented bundles. Additionally, the study found a substantial amount of LBs (between 20% and 25%) that were shared between the two corpora.

Finally, Pérez-Llantada (2014) compared LBs across three language varieties of expert academic writing (L1 English, L2 English written by Spanish speakers, and L1 Spanish). After analyzing the structures and functions of bundles specific to one or two language variables, she argued that the use of LBs by L2 writers deviated from L1 norms and concluded that L2 expert writers' formulaicity was 'hybrid' —largely, but not completely, native-like (Pérez-Llantada 2014: 93).

Additional studies on the L1 influence in L2 academic writing offered further insights into the processes behind the phenomenon. Rica Peromingo (2012) investigated L1 transfer in argumentative essays by Spanish learners of English. In particular, the study looked at linking adverbial LBs that create textual cohesion (e.g., *in other words*). The learners in the study demonstrated overuse of L2 English adverbials that had a similar meaning to those used in Spanish (e.g., *in conclusion = en conclusión*). Rica Peromingo hypothesized that the structural and semantic similarity of the LBs could explain the observed transfer. L1 transfer in learners' production of LBs that are semantically and structurally similar in learners' L1 and target L2 was also supported by Allen (2011). The study suggested that the overuse of certain LBs (e.g., *it can be said (that)*) in final course research papers written by Japanese learners of English might occur due to the proximity of these bundles to similar L1 Japanese bundles. Allen (2011: 119) attributed this transfer pattern to lexical priming in one's L1 that may facilitate writing in an L2.

While the studies above have provided some evidence for possible L1 transfer in the use of LBs, this evidence is based solely on the finding that a certain construction found in L2 writing exists in learners' L1. Paquot (2013) argued that such an approach may be problematic as it involves post-hoc guessing on the side of the researcher. In order to address this issue, she examined the effects of transfer on French EFL learners' use of LBs applying Jarvis's (2000) framework for the study of L1 transfer that consists of three potential sources of transfer evidence (see Section 2.3 below). Conducting a LB analysis on the French part of the *International Corpus of Learner English* (ICLE),[1] Paquot (2013) detected that learners' application of three-word LBs in writing was associated with lexico-grammatical as well as functional frequency patterns in French. Based on these results, Paquot argued that the first language of learners may prompt them to use LBs in a way that is not typical for English. In a follow-up study, Paquot (2017) investigated the

---

[1] https://uclouvain.be/en/research-institutes/ilc/cecl/icle.html

preferred use of LBs expanding the analysis in the writing of French and Spanish learners of English. Using the frequency data, Paquot found strong positive correlations between the frequency of discourse organizational and stance-oriented LBs in learners' written production and its equivalent form in the learners' L1. Making use of the same framework, Güngör and Uysal (2020) recently investigated the cross-linguistic influence of L1 Turkish on L2 English on the learners' production of four-word LBs. The study revealed that 45 percent of bundles in L2 English writing were distinctive to Turkish authors.

Taken together, previous studies pointed out deviations in learners' use of LBs. Some have compared L1 and L2 LBs and argued that learners, irrespective of their L2 proficiency levels, misuse the formulaic sequences in L2 English academic writing (e.g., Chen and Baker 2010; Salazar 2011; Ädel and Erman 2012; Esfandiari and Barbary 2017). Although these studies claimed that the misuse of LBs in L2 texts might be due to the L1 transfer, they oftentimes assumed L1 interference just based on the analysis of the L2 texts without analyzing the data in L1 (Gilquin and Paquot 2008). At the same time, those studies that included learners' L1 as another point of comparison (e.g., Pérez-Llantada 2014) have disregarded the importance of evidence that is rooted in established frameworks. Lastly, the studies that made use of such frameworks are limited to certain L1s and need to be expanded to learners from other L1 backgrounds.

2.3. L1 influence identification framework

As argued in the previous section, few studies that examined L1 transfer evidence in L2 learners' production of LBs in academic writing grounded their investigations in transfer frameworks. To this end, Paquot (2013) adapts Jarvis's (2000) framework for assessing L1 transfer. According to Paquot (2013: 393−394), the framework requires three types of comparisons to be considered by studies in order for transfer to be supported by sufficient evidence: (1) intra-L1-group homogeneity in learners' L2 performance where learners that share an L1 display similar patterns of use of a specific L2 feature; (2) inter-L1-group heterogeneity in learners' L2 performance where learners from different L1s do not share the same patterns; and (3) intra-L1-group congruity between learners' L1 and L2 performance where the comparison of learners' use of a feature in their L1 and L2 reveals similarities. In her later study, Paquot (2017), referring to Jarvis (2000: 258), emphasized that intra-L1-group congruity is the strongest type of evidence for L1 influence, as the comparison of learners'

L1 and L2 production can demonstrate L1 features that motivate patterns of use of similar features in learners' L2. Additionally, intra-L1-group congruity lends itself to a statistical approach to L1 transfer examination, which is crucial in Jarvis's framework.

3. CORPUS AND METHODOLOGY

The corpora examined in this study were comprised of research articles in the field of Educational Psychology. These articles were written by L1 English (PSY-ENG1), L2 English (PSY-ENG2), and L1 Russian (PSY-RUS1) expert writers. It is important to remember that for the sake of comparability, all of the L2 English articles were written by Russian native speakers (see below). The articles came from three major peer-reviewed journals in the field of psychology: *American Psychologist* (L1 English), *Psychology in Russia* (L2 English), and *Национальный Психологический Журнал (Nacionalniy Psihologicheskiy Zhurnal)* (L1 Russian). *American Psychologist* was chosen on the basis of its high impact factor (4.856) and the fact that it is the official journal of the *American Psychological Association*. Since impact factor is not calculated for Russian psychological journals, the other two periodicals were selected because they are published by the leading research universities in Russia. Overall, the corpora in this study were designed for contrastive descriptive research of LBs in written discourse of L1 English, L2 English, and L1 Russian academic professionals and, therefore, were made comparable with regard to register, discipline, communicative purposes, and authors' level of expertise.

One concern that emerged during the first stages of data collection was determining the first language of a writer. Following Pan *et al*. (2016: 63), L1 Russian (and thus L2 English) writers were defined as authors affiliated to an institution located in a country where Russian was spoken as the first language. Additionally, the author's first and last names had to be considered native to these countries. Articles by writers with arguable names were excluded from the corpora. The same procedure was implemented to identify L1 English writers. The final corpus structure is shown in Table 1.

	PSY-ENG1 (L1 English)	PSY-ENG2 (L2 English)	PSY-RUS1 (L1 Russian)
Number of texts	61	85	91
Average number of words per text	6,730.50	4,842.80	4,525.10
Total number of words	410,558	411,637	411,787
Total number of types	19,025	17,149	53,399
TTR	5.26	4.77	12.97
Standardized TTR	5.09	4.74	12.53

Table 1: Summary of built corpora

The process of corpus building for this study consisted of two steps. During the first step of data collection, articles published between 2017 and 2019 were downloaded for each corpus. Importantly, only research articles, descriptions or research methodology, and literature reviews were included in the corpora; that is, other types of texts published in the journals (e.g., editor's notes, reviews, opinions) were excluded from the analyses. After the extraction, all articles were cleaned of meta-data and references as well as text in languages other than the target ones. For example, if an L1 Russian article contained text in a language other than Russian, this text was removed from the article before its inclusion in the corpus. In order to match the corpora on the number of words, additional articles from 2015 and 2016 were downloaded from the journals in the PSY-ENG2 and PSY-RUS1 corpora. This resulted in three corpora with the same number of words, although slightly different text counts (see Table 1).

3.1. Identification of lexical bundles

In order to retrieve the frequency lists of bundles and compute tokens and types of LBs from the collected corpora, the study used the *Natural Language Toolkit* (NLTK) library of *Python* (Bird *et al*. 2009) and followed the LB extraction steps outlined in Ren (2021). Log-likelihood values were calculated in *R,* a free statistical environment (R Core Team 2019) and compared to establish whether the frequency of the bundles used only by L1–L2 English and the frequency of the bundles used only by L2 English–L1 Russian writers differed significantly. Significant differences in the use of similar LBs between L1 and L2 English corpora would indicate that Russian learners of English demonstrate professional writing that is different from L1 English writing. Conversely, lack of significance in the use of similar LBs between L2 English and L1 Russian corpora would suggest L1 transfer in the writing of Russian authors in English.

Three criteria were considered in the identification of LBs: bundle length, frequency, and dispersion. The study focused on four-word bundles for PSY-ENG1 and PSY-ENG2 to make the analysis more manageable and comparable to those of other studies (e.g., Chen and Baker 2010; Pérez-Llantada 2014; Pan *et al.* 2016). Moreover, this length seems to display a wider variety of structures and functions for analysis than three- and five-word bundles (Cortes 2004; Hyland 2008). Cortes (2004: 401) also noted that three-word bundles are often embedded in four-word bundles (e.g., *at the end* and *at the end of*). However, it was deemed necessary to also include three-word LBs in the process of retrieval and analysis of LBs in the PSY-RUS1 corpus. The Russian language has a rich and highly inflectional morphological system. Importantly, inflectional morphemes embedded in a word can indicate tense, voice, and number (cf. *on the other hand* vs. *с другой стороны* [*s drugoy storony*]). Moreover, some functional words, for example, definite and indefinite articles do not exist in Russian. Therefore, it is often the case that a four-word bundle in English has a three-word equivalent in Russian (*the table shows that* vs. *таблица показывает что* [*tablitsa pokazyvaet chto*]). Thus, both three-word and four-word LBs were analyzed from the PSY-RUS1 corpus.

As for the criterion of LB frequency, recent studies made use of varied thresholds ranging between 20–40 times per million words (e.g., Biber *et al*. 2004; Hyland 2008; Chen and Baker 2010). For this study, a high cut-off of 40 per million was set. This threshold is helpful in filtering out content bundles as well as bundles containing discipline-specific nouns (Ädel and Erman 2012; Pérez-Llantada 2014). The dispersion criterion for this study was set at 10 percent. This means that a lexical bundle had to appear in at least 10 percent of the texts in a corpus to be considered for inclusion in the analysis. Previously, researchers have chosen different dispersion criteria for their studies varying between three to five texts in a corpus (Biber and Conrad 1999; Chen and Baker 2010; Ädel and Erman 2012). Pan *et al*. (2016), for example, established a LB dispersion threshold of five texts for an 87-text corpus (5.7%) and ten texts for a 179-text corpus (5.6%). Although this approach is effective for comparing corpora with the same number of texts, it can present a methodological problem if the corpora are not matched for this number (Hyland 2008). Setting a percentage dispersion threshold was especially important for the second step of lexical bundle extraction in this study since the three corpora differed in the number of texts (see Table 1). The established dispersion threshold was also considered adequate given the previous practices.

LBs for the analysis were identified on the basis of their word forms and not lemmas. In other words, inflected variants of the same lemma were treated independently. This decision was especially important in the case of the PSY-RUS1 corpus since, as mentioned above, Russian has a highly inflectional morphology, and the identification of LBs based on lemmas might have caused loss of important comparison points between the corpora. The retrieved bundles were checked manually for the remaining area-specific content bundles. Content bundles involving proper nouns (*American Psychological Association*) were excluded and the bundles related to conducting research in general (e.g., *majority of the informants*) were kept. Following Chen and Baker (2010:33), the overlapping bundles in the PSY-RUS1 list were merged; thus, three-word bundles that were parts of four-word bundles in the list and occurred with the exact same dispersion and frequency were merged. For example, the three-word bundle *то же время* (*to zhe vremya*) 'the same time' appeared in 23 texts and had a frequency of 150 words per million. A similar four-word bundle *в то же время* (*v to zhe vremya*) 'at the same time' has the same dispersion and frequency. Therefore, the two overlapping bundles are combined into *(в) то же время+* (*(v) to zhe vremya+*) / '(at) the same time+' in the final list. The merged bundles are indicated with a plus (+) sign in the compete lists provided in Appendix 1.

3.2. Application of Jarvis's (2000) framework for additional L1 transfer evidence

As mentioned above, to provide further statistical evidence of L1 influence on Russian L2 English writers' production of LBs, the study used the L1 transfer assessment framework proposed by Jarvis (2000). Following Paquot's (2017: 6) claim that the intra-L1-group congruity between learners' L1 and L2 performance presents the strongest type of evidence for L1 influence (also Jarvis 2000: 258) and for the sake of feasibility, the present study made use of this effect to further examine L1 transfer in Russian writers' LB use in L2 English writing. As Paquot (2013: 400) notes, the simplest way to test the intra-L1-group congruity criterion is to check whether there are bundles that are shared between learners' L1 and L2 writing. Thus, frequent LBs in PSY-ENG2 and PSY-RUS1 were compared for the presence of overlapping bundles.

3.3. Translation and analysis of bundles

To single out the bundles shared in the three language varieties as well as bundles shared by only PSY-ENG2 and PSY-RUS1 (Jarvis's (2000) intra-L1-group congruity), the L1 Russian LBs were translated into English by two researchers (the author and another applied linguistics scholar) whose native language was Russian and who had done similar translation work before. The translations from Russian to English were done with the help of the *Collins Russian-English Dictionary*.[2] Importantly, the translations were maintained as close as possible to the original. In other words, the researchers aimed at word-for-word translations; however, in cases where it was not possible, a lexical bundle with the most similar meaning was used. The translations provided by both researchers were compared in order to ensure the validity of the English equivalents for the Russian LBs. All discrepancies were discussed and resolved reaching 100 percent agreement between the two translators. The three LB lists were then compared manually. Log-likelihood analyses were performed with the bundles shared between PSY-ENG1 and PSY-ENG2, as well as between PSY-ENG2 and PSY-RUS1, to find out the significant differences in bundle frequencies in these language varieties. LBs unique to only one corpus were also identified.

After the quantitative analysis regarding the LBs extracted from the three corpora, the 50 most frequent bundles in each list were classified. Biber *et al*.'s (2004) framework (modified by Hyland 2008 and Pan *et al*. 2016) was used to compare the LBs based on their discourse functions. LBs were classified into three major categories: research-oriented (parallel to 'referential' bundles in Biber's *et al*. (2004) framework), text-oriented (parallel to 'discourse-organizing'), and stance-oriented bundles. Bundles identified as research-oriented were those that explained the procedures in a study as well as its structure (e.g., *at the same time*). Text-oriented bundles (e.g., *in addition to*) were those involved in organization of the text of an article and its argumentative elements. Finally, stance-oriented bundles (e.g., *it is possible that*) had the function of conveying an author's evaluation and attitude towards the reported information. The bundles were first classified by two raters trained in the field of corpus linguistics and familiar with the framework. The initial agreement rate between the raters was 82 percent. After an inter-rater norming session was held, disagreements in functional identification of LBs were resolved resulting in 100 percent agreement. As the final step of the functional analysis,

---

[2] https://dictionary.reverso.net/russian-english/

Chi-square tests were also performed to check for significant differences in the functional distribution of bundle types in the three corpora.

4. RESULTS AND DISCUSSION

The established frequency and dispersion cut-offs resulted in 82 bundles identified in PSY-ENG1, 223 bundles in PSY-ENG2, and 264 bundles in PSY-RUS1. Appendix 1 provides a complete list of the extracted LBs with their frequencies normalized per million words (pmw). Overall, the amount of LBs retrieved from the three corpora supports the view that the academic written register can be clearly characterized by formulaicity and fixedness of expressions (Pérez-Llantada 2014). If compared to previous research in the area of LBs in academic writing, Cortes (2004) reported 54 frequent bundles in her corpus of writing in history and 109 bundles in biology writing. Pérez-Llantada (2014) was able to retrieve a total of 56 bundles in L1 English, 77 in L2 English, and 114 in L1 Spanish writing. With regard to the total number of LBs in the three corpora, L1 English writing displayed the lowest amount of frequent LBs (83), especially since both L2 English and L1 Russian writing contained more than twice the amount of bundles (227 and 264). A similar trend was displayed in Hyland (2008) and Römer (2009) with L2 writers producing a larger number of bundles than L1 English writers. This finding offers support to Ellis (2002) who suggested that L2 production is oftentimes more formulaic than L1 production. Additionally, the finding also seems to support the hypothesis expressed by Pérez-Llantada (2014), who suggests that an observed wider range of bundles can be interpreted in terms of lexical variety of a given language. Thus, the fact that PSY-RUS1 showed the highest total number of word types (53,399), Type-Token Ratio (TTR) (12.97), and Standardized Type-token Ratio (STTR) (12.53) compared to PSY-ENG1 and PSY-ENG2, as indicated in Table 1, could be viewed as indirect evidence for the lexical richness of the Russian language and, consequently, the higher number of the extracted LBs. However, this hypothesis does not explain the large number of bundles in PSY-ENG2 with the word types, TTR, and STTR being close to PSY-ENG1. Another explanation for the differing numbers of frequent LBs can be the possibility of L1 influence in writing (Paquot 2014). Russian learners of English might be adapting some of the LBs from their native language into L2 English writing. Finally, it is also possible that because the PSY-ENG1 corpus included a smaller number of texts,

it yielded fewer bundles despite the same dispersion cut-off (see Chen and Baker 2010: 43).

4.1. Core bundles

To identify the bundles that were shared between all three corpora, the extracted LBs were compared manually. A total of six bundles were shared between three corpora, representing 7.3 percent of the bundles in L1 English writing, and three percent in L2 English writing as well as in L1 Russian writing (see Table 2). It can be assumed that these core bundles are extremely useful in both English and Russian for various discourse purposes. Supporting Pan *et al*. (2016: 68), the majority of these core bundles serve the text-organizing function (*at the same time*, *as well as the*, *in the case of*), with two bundles functioning as research-organizers (*at the end of*, *is one of the*) and one bundle having a stance function (*it is important to*).

Interestingly, some of these bundles had differing normalized frequencies; for instance, *at the same time* was the most frequently occurring LB in the PSY-ENG2 corpus (303 pmw), but barely met the threshold in the PSY-ENG1 corpus (20 pmw). In contrast, *it is important to* appeared 118 times per million words in PSY-ENG1 and only 32 times in PSY-RUS1.

The use of the core LBs in L2 professional writing might extend on more than just the two languages under analysis. After comparing the core LBs in our study to those in Chen and Baker (2010) and Ädel and Erman (2012), five out of six bundles overlapped in the two studies. The only exception was *at the end of*, which was identified as a shared lexical bundle only in Ädel and Erman (2012). Recall that both studies compared L1 English academic writing to learner writing in by native speakers of other languages (Swedish and Chinese). It seems, therefore, that these core bundles are acquired by L2 English writers with different L1 backgrounds and are not indicative of L1 transfer.

Lexical bundle	Frequency, pmw		
	PSY-ENG1	PSY-ENG2	PSY-RUS1
it is important to	118	75	32
as well as the	60	170	85
at the end of	38	63	23
in the case of	25	168	32
is one of the	24	113	49
at the same time	20	303	150

Table 2: Bundles shared by all three corpora

4.2. Bundles shared in L1 English and L2 English

A total of 17 bundles were found to overlap in L1 and L2 English writing, representing 20.3 percent of the L1 English writing and 7.5 percent of the L2 English writing. If we add these bundles to the core bundles shown in Table 3, PSY-ENG1 and PSY-ENG2 share a total of 23 LBs (28% and 10.1% of L1 and L2 writing respectively). It appears that this amount of overlap in bundles is quite large, especially in comparison to the results by Chen and Baker (2010), who found 16 percent of LBs overlapping between L1 and L2 English writing.

Lexical bundle	Frequency, pmw		
	PSY-ENG1		PSY-ENG2
in the context of	135		142
it is important to	**118***		75
as well as the	60	**170***	
one of the most	48	**97** *	
in the form of	41		72
at the end of	38	**63***	
in the development of	38		35
it is possible that	38		38
with respect to the	38		35
in addition to the	33		28
the nature of the	33		28
as a result of	31	**72***	
the context of the	31		38
in terms of the	25		28
in the case of	25	**168***	
it is possible to	24	**69***	
is one of the	24	**113***	
and the development of	21		35
it should be noted (that)+	21	**91***	
the importance of the	21		22
at the time of	20		28
for the development of	20	**85***	
at the same time	20	**303***	

* = significant at *p*<0.05

Table 3: Bundles shared only between PSY-ENG1 and PSY-ENG2

Following Simpson-Vlach and Ellis (2010), log-likelihood values were calculated for the overlapping bundles. The list of overlapping bundles is presented in Table 3 together with the results of the log-likelihood analysis with the core bundles underlined and the numbers in bold indicating overuse. The log-likelihood statistics indicate that L2 English writing displays an overuse of some of the shared bundles (e.g., *as well as the*, *in the case of*, *at the same time)* including all of the core bundles. Similar findings were reported by Ädel and Erman (2012) who found that shared LBs were overused in L2 writing. It may be the case that L2 writers are more familiar with these bundles and feel confident using them in writing (Granger and Rayson 1998; Pérez-Llantada 2014). Ellis (2008) also suggests that L2 writers might have memorized these LBs and routinized them in their writing. Only one bundle (*it is important to*) was underused in the PSY-ENG2 corpus. This underuse may be due to the fact that Russian academic writers tend to use fewer stance bundles, as illustrated in the functional analysis in Section 4.5 below, pointing at possible L1 transfer.

Compared to the complete PSY-ENG1 and PSY-ENG2 lexical bundle lists, the data seem to support Swales's (2005: 10) and Ädel and Erman's (2012) observation that attended *this*-bundles with the meta-discursive head nouns (*of this study is*, *this point of view*, *the results of this study*) are more common in non-native writing.

4.3. Bundles shared in L2 English and L1 Russian (intra-L1-group congruity)

To further investigate the L1 transfer evidence within Jarvis's (2000) framework, the extracted lexical bundle lists were compared to find bundles that overlapped in PSY-ENG2 and PSY-RUS1 (see Table 4). A total of 22 bundles were shared between PSY-ENG2 and PSY-RUS1 corpora, which comprise 9.7 percent of the frequent bundles in L2 English writing and 8.3 percent in L1 Russian writing. If merged with the core bundles, there is a total of 28 bundles shared between the two corpora (12.3% and 10.6% in PSY-ENG2 and PSY-RUS1, respectively). The overlapping LBs between L2 English writing produced by Russians and L1 Russian writing further suggest the possibility of L1 influence. Yet, the significant log-likelihood values of the overlapping bundles presented in Table 3 indicate, in accordance with Pérez-Llantada's (2014) findings, that very few bundles in L2 English writing are used in a Russian native-like manner. This suggests that L2 writers' usage of bundles is not fully native-like and represents a combination of both L1 English and L1 Russian academic writing. At the same time, bundles like *in the*

present study / в данном исследовании (*v dannom issledovanii*), *we can say that / мы можем сказать что* (*my mozhem skazat chto*), or *an important role in / важную роль в* (*vazhnuyu rol v*) do not significantly differ in their use in L2 English and L1 Russian pointing at the possible L1 transfer, especially since these bundles do not occur in the corpus of L1 English writing (see Appendix 1). Even those LBs that are used in L1 Russian writing significantly more often than in L2 English writing (e.g., *on the one hand / с одной стороны* [*s odnoy storony*], *in this case the / в этом случае* [*v etom sluchaye*], *and on the other / а с другой* [*a s drugoy*]) are potentially indicative of cross-linguistic transfer as they do not appear in L1 English writing at all.

Some other important observations about L1 Russian LBs emerged after examining the lists more closely. Similarly to English, a lot of four-word Russian bundles had three-word bundles embedded in them (e.g., *свидетельствуют о том (что)+* [*svidetelstvuyut o tom (chto)+*] / *indicate (that)*, *вывод о том (что)+* [*vyvod o tom (chto)+*] / *conclusion about that (that)+*, *несмотря на то (что)+* [*nesmotrya na to (chto)+*] / *despite the fact (that)+*, *так же как (и)+* [*tak zhe kak (i)+*] / *same as (and)*, *(с) нашей точки зрения+* [*(s) nashey tochki zreniya+*] / *(from) our point of view*). It seems that this embedding is dictated by the syntactic structure of the language: the words in brackets in the examples are prepositions and conjunctions that are, in most cases, required by the words they follow or precede.

Bundles like *вывод о том (что)+* (*vyvod o tom (chto)+*) / *conclusion about that (that)+* deserve special attention in this study. In this bundle, the demonstrative pronoun *том* (*tom*) ('that' in prepositional case) acts as the head noun in the noun phrase of the prepositional phrase *о том* (*o tom*; 'about that'). This prepositional phrase can be roughly translated as 'about the fact' (*о том факте* [*o tom factye*]) with the Russian version being an acceptable and widely used phrase. However, the noun *fact* is often omitted in Russian because it is contextually predictable and, therefore, redundant (Jaeger and Tily 2011: 328). In PSY-RUS1, 22 out of 231 bundles (9.5%) had a similar structure with the pronoun *то/том* (*to/tom*; 'that'/'that' in prepositional case) taking the place of the noun. Interestingly, there were seven bundles in PSY-ENG2 that contained the word *fact* (*the fact that the*, *to the fact that*, *by the fact that*, *due to the fact that*, *in the fact that*, *of the fact that*, *explained by the fact*) and none in PSY-ENG1. This finding serves as another indicator that L1 transfer may be happening in L2 writing. It is noteworthy that apart from the 22 shared bundles there were cases when the bundles closely resembled each other in

PSY-ENG2 and PSY-RUS1 LB lists. For example, the PSY-ENG2 bundle *in the present study* did not occur in PSY-RUS1, but it is very similar to an L1 Russian bundle *в данной статье* (*v dannoy statye* 'in the present article'). Corresponding cases include L2 English bundles *the study showed that*, *we assume that*, *in his opinion* that have close equivalents in L1 Russian writing.

The comparison of PSY-ENG2 and PSY-RUS1 bundles also provided further methodological considerations with regard to the length of *n*-grams in English and Russian. It has been argued above that a four-gram is the most commonly studied length of *n*-grams in most studies on lexical bundles. However, the present analysis revealed that the length of similar *n*-grams in Russian and English often does not match. When the retrieved three- and four-grams in PSY-RUS1 were translated into English, their length changed; many three-grams in Russian became one-grams in English (*в том числе* [*v tom chisle*] *including*, *в настоящее время* [*v nastoyashee vremya*] */ currently*, *тем не менее* [*tem ne menee*] *nevertheless*, *включает в себя* [*vklyuchayut v sebya*] *includes*, *на сегодняшний день* [*na segodnyashniy den*] */ nowadays*, *по всей видимости* [*po vsey vidimosti*] */ evidently*, *в последнее время* [*v poslednee vremya*] *recently*). A few bundles also became longer after being translated from Russian into English (*важно отметить что* [*vazhno otmetit chto*] */ it is important to note that*, *следует подчеркнуть что* [*sleduet podcherknut chto*] */ it should be emphasized that*, *это связано с* [*eto svyazano s*] */ it is connected to*). Therefore, it is evident that a larger range of LB lengths needs to be included in cross-linguistic bundle studies, especially when one of the compared languages is so morphologically rich, as is the case with Russian.

Lexical bundle	Frequency, pmw		
	PSY-ENG2	PSY-RUS1	
at the same time / в то же время	**303***		150
a high level of / высокий уровень того	**224***		39
as well as the / так же как и	**170***		85
in the case of / в том случае	**168***		32
on the other hand / с другой стороны	**145***		105
in the process of / в процессе того	**142***		39
is one of the / является одним из	**113***		49
with a high level / с высоким уровнем	**110***		39
with the help of / с помощью того	**101***		29
on the one hand / с одной стороны	94	**153***	
it is important to / является важным то	**75***		32
at the end of / в конце того	**63***		23
are presented in table / представлены в таблице	**56***		35
in this case the / в этом случае	50	**91***	
as well as to / так же как и чтобы	38		39
(it) can be assumed that + / можно предположить что	38	**91***	
in the present study / в данном исследовании	38		39
(at) the same time they + / в то же время они	38		23
as well as in / так же как и в	35	**85***	
in the fact that / в том что	32		37
we can say that / мы можем сказать что	32		26
an important role in / важную роль в	28		26
as well as a / так же как и	28		23
and on the other / а с другой	25	**42***	
as well as their / так же как и их	25		20
not only in the / не только в	20	**55***	
to a lesser extent / в меньшей степени	20	**35***	

* = significant at *p*<0.05

Table 4: Bundles shared only between PSY-ENG2 and PSY-RUS1

4.4. Functional classification

To answer the second question in the study, the first 50 bundles in the three lists were classified according to their discourse function in the articles. The complete analysis of the first 50 bundles can be found in Appendix 2. As seen in Figure 1, PSY-ENG1 and PSY-ENG2 display similar proportions of the three main functional categories. Research-oriented bundles constitute the largest category in both corpora, with 42 percent and 60 percent respectively, whereas stance bundles comprise 22 percent and 12 percent of the 50 most frequent LBs in the two corpora. Similarly, Pérez-Llantada (2014) found that the bundles shared by L1 and L2 English most commonly perform a referential function. Turning to PSY-RUS1 bundles, text-oriented LBs clearly rank as the largest category with 72 percent, followed by research-oriented bundles (18%) and stance bundles (10%). This LB distribution partially supports Pan *et al.* (2016), who also found stance to be the smallest functional category in L1 Chinese writing; however, the text-oriented category

was the largest one in L1 and L2 English writing in contrast to what happens in the current study, which shows the dominance of research-oriented bundles in L1 and L2 English.
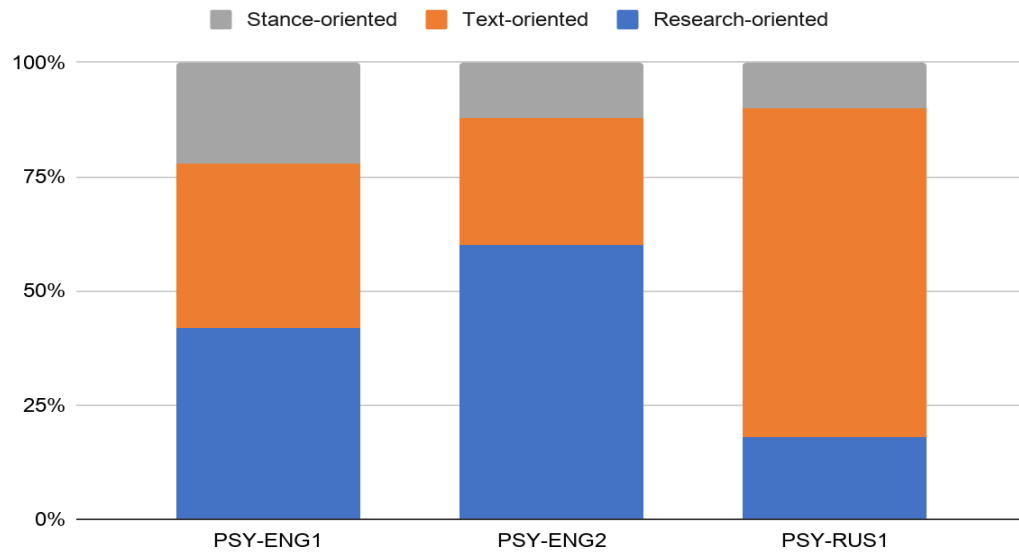

Figure 1: Functional distribution of the 50 most frequent bundles in PSY-ENG1, PSY-ENG2, and PSY-RUS1

Table 5 presents the results of a Chi-square test showing a significant medium-sized difference in the functional distribution of the bundles between the three corpora ($\chi2=$ 22.71, 4 df, Cramer's V = 0.272, $p<0.05$).

		Research-Oriented	Function Text-oriented	Stance-oriented	Total
PSY-ENG1	Count	21	18	11	50
	Expected Count	20	22.7	7.3	50
	Adjusted Residual	0.4	-1.6	1.8	
	Probability value	0.4839	0.1615	0.3681	
PSY-ENG2	Count	30	14	6	50
	Expected Count	20	22.7	7.3	50
	Adjusted Residual	3.5	-3	-0.7	
	Probability value	0.0005	0.0027	0.4839	
PSY-RUS1	Count	9	37	4	50
	Expected Count	20	22.7	7.3	50
	Adjusted Residual	-3.9	4.6	-1.1	
	Probability value	0.00009	0.000004	0.2713	

Table 5: Results of the Chi-square test of the functional distribution of the 50 most frequent bundles in PSY-ENG1, PSY-ENG2, and PSY-RUS1

To find out where exactly the significance lies, post-hoc tests were conducted, and adjusted residuals and probability values were calculated and compared to the

Bonferroni-adjusted *p*<0.005. As revealed in Table 5, the distribution of functions was not significantly different from the expected counts in PSY-ENG1 bundles; however, research-oriented and text-oriented bundles were found significantly more frequent in L1 Russian writing (*p*<0.005). It is noteworthy that the results of the Chi-square test for these two functions were also significant with regard to L2 English bundles, although the distribution was the opposite. There were significantly fewer text-oriented bundles and more research-oriented bundles than expected.

Looking closely at the functional subcategories of the bundles (available in Appendix 2), we can notice that, in line with Hyland (2008), research-oriented bundles in the three corpora are represented by the following subcategories: description, location, quantification, procedure, and topic. It seems that L1 English and L2 English professional writers make use of description bundles more often, focusing on providing identification of new information for the readers (Biber 2009). Research-oriented bundles in L1 Russian writing are remarkably less common. However, within the subcategories of text-oriented bundles, there is a prevalence of transition signals in L1 Russian writing with 72.4 percent of all the bundles in this category. This subcategory is also the largest one in L2 English writing (42.6%). The main function of text-oriented bundles is to establish textual cohesion through signaling transition or discussion of results, framing the discussion, and guiding the reader through the overall structure of the article. In other words, these bundles can be described as meta-discourse (Ädel and Erman 2012). It has been previously reported (Ädel 2006) that L2 learners tend to overuse meta-discourse in academic writing; however, this is not the case in my data, perhaps due to a higher L2 proficiency of expert writers.

In contrast, framing signals are the most prominent subcategory in L1 English writing. Framing is the only subcategory in text-oriented bundles where L2 writers use bundle tokens significantly less frequently than L1 writers do (Pan *et al*. 2016). With regard to this subcategory, it is interesting to note that two out of five LBs used by L2 English writers overlap with L1 English writing (*in the context of* and *in the case of*). Römer (2009) and Chen and Baker (2010) noticed that the bundle *in the context of* was rarely used by novice L2 learners; however, it is highly frequent in the PSY-ENG2 corpus (146 times pmw). This, again, may indicate that the use of LBs becomes more native-like with the growing proficiency of L2 professional writers. Additionally, the high frequency

of the bundle might have occurred due to writers' L1 influence, although an equivalent bundle was not detected in the PSY-RUS1 corpus.

Stance features and engagement features were used to convey the author's interpretation in professional writing, but the proportions were somewhat small, as noted in previous research (Biber *et al*. 2004; Hyland 2008; Chen and Baker 2010). Although the distribution of stance-oriented bundles was not found significantly different from the expected counts, it is still noteworthy that the PSY-ENG1 list contained almost twice as many stance bundles as PSY-ENG2 and PSY-RUS1. Similar observations about the lack of control of formulaic language expressing stance in L2 professional writing were made by previous studies (Ellis 2008; Granger and Meunier 2008; Pérez-Llantada 2014). This lack of control may be attributed to L1 syntactic and lexical transfer. If we compare the stance-oriented bundles in L2 English and L1 Russian writing, several similarities emerge, the most outstanding being the use of quite a direct noun *fact* (*the fact that the*, *to the fact that*, and *тот факт что* [*tot fakt chto*] / *the fact that*). It seems, therefore, that the stance feature bundles in L2 English and L1 Russian writing might not display enough hedging. Pérez-Llantada (2014) hypothesized that the paucity of stance meanings that builds a potentially face-threatening discourse can be attributed to the mismatch of L1 pragmatic norms. Pragmatic mismatches have also been reported in Philippine scholars (Salazar 2011: 193) and in Finnish undergraduates who show less variation in stance bundles than their L1 English counterparts (Ädel and Erman 2012). As explained in Granger (1998) and Chen and Baker (2010), the L2 English writers use fewer hedges because they have not acquired full pragmatic competence yet. At the same time, the presence of overlapping stance bundles in PSY-ENG1 and PSY-ENG2 (*it is important to*, *it is possible that*) points at a developing proficiency in L2 English writing and suggests that the use of stance bundles in L2 English writing is influenced by both L1 English and L1 Russian distribution of LBs.

5. CONCLUSION AND FUTURE DIRECTIONS

The present study explored the use of LBs in L1 and L2 professional writing in the field of Educational Psychology. In particular, the study investigated the nature and functions of LBs in L1 and L2 English, as well as L1 Russian articles, in an effort to examine the similarities between L1 English and L2 English writing and detect possible evidence for

cross-linguistic transfer between L1 Russian and L2 English writing produced by Russian speakers.

Regarding the first research question that centered around the frequency evidence of cross-linguistic transfer, the results indicated that Russian authors display some evidence of L1 transfer in their L2 English writing (Bybee 2008; Paquot 2014). Specifically, the intra-L1-group congruity evidence collected within the framework proposed by Jarvis (2000) showed that a number of identified bundles were shared between L2 English and L1 Russian writing and did not occur in L1 English writing. A similar trend was uncovered in Güngör and Uysal (2020), where bundles specific to Turkish learners of English constituted almost 50 percent of the LB list. Further evidence of transfer was found in the functional analysis of LBs. That is, the high-frequency bundles in L2 English and L1 Russian writing included fewer stance-oriented LBs than in L1 English. Additionally, within text-oriented bundles, transition signals were the largest subgroup proportionally compared to L1 English bundles, where framing was the most common function of text-oriented LBs. On the other hand, L1 and L2 English writing demonstrated similar distribution of functions overall with research-oriented bundles being the largest category, while text-oriented bundles were the most common in Russian. Finally, my analysis revealed a list of core bundles that were shared among L1 and L2 English speakers. Thus, the study offered some evidence for cross-linguistic transfer in English writing produced by Russian authors, although it was not pervasive in the analysis of the extracted LBs and their functions.

The corpus-driven approach of the study supported the current research in LBs, showing that formulaic sequences are a fundamental feature of the academic register across language variables. However, the number of frequent LBs was found higher in PSY-ENG2 and PSY-RUS1 writing in comparison to PSY-ENG1. This result disconfirms previous research (e.g., Chen and Baker 2010; Ädel and Erman 2012), which found that non-native speakers possess a more restricted inventory of bundles than native speakers. Thus, the present study contributes to a unique strand of research (cf. Pérez-Llantada 2014) that uses corpus evidence to demonstrate that the L2 English writing reflects a 'hybrid' nature of formulaic language. In this study, L2 English displays a small number of register-determined bundles also shared by L1 English and L1 Russian. At the same time, it also includes a considerable percentage of formulaic sequences used by the L1 English writers as well as bundles transferred from L1 Russian. Furthermore, through

the functional analysis of the most frequent LBs it was found that both L2 English and L1 Russian employ fewer stance-oriented bundles, and the number of text-oriented bundles is closer between L1 and L2 English writing. Finally, one cannot forget about the case of *fact* in PSY-RUS1 writing that seems to influence the composition of LBs in PSY-ENG2. In brief, L2 English professional writing is partly, but not fully, native-like, possibly due to cross-linguistic influences from the writers' L1.

The present exploratory study poses several directions for future research. To control for content-specific bundles, the study only focused on one discipline. However, research with monolingual corpora has empirically confirmed the existence of 'discipline-sensitive' bundles in the context of research article writing (e.g., Cortes 2004; Hyland 2008). It would be worth conducting interlinguistic comparison of bundles across the disciplines to determine what bundles are specific to those disciplines and what discourse functions these bundles perform in L1 and L2 writing. It would also be of theoretical interest to further investigate the hybrid formulaic nature of L2 English research articles in languages other than Russian and Spanish (Pérez-Llantada 2014). With regard to methodology, another limitation of the current study has been the absence of a L2 English corpus that was produced by learners with the L1 background different from Russian. While the study was able to make use of previous comparable research that identified LBs in order to meet one of the criteria in Jarvis's (2000) framework (intra-L1-group congruity), a corpus built specifically for the study would facilitate a more fine-grained search of the bundles that could be shared between L2 English learners from different backgrounds and thus contribute to our understanding of L1 transfer in Russian by providing the other two types of evidence from Jarvis (2000). It also needs to be stressed that only four-word bundles were considered in PSY-ENG1 and PSY-ENG2. The process of translation of Russian bundles into English showed that some of the translated LBs did not match in length to the original. It was often the case that a three-word Russian bundles could be translated as one word in English. Thus, a fuller picture of the use of formulaic language across the corpora could have been given if more bundle lengths had been included. Finally, as pointed out by one of the reviewers, comparative analyses of translations are inherently problematic as not all LBs have exact equivalents between languages.

REFERENCES

Ädel, Annelie. 2006. *Metadiscourse in L1 and L2 English.* Amsterdam: John Benjamins.

Ädel, Annelie and Britt Erman. 2012. Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes* 31/2: 81–92.

Allen, David. 2011. Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education* 1: 105–127.

Biber, Douglas. 2009. A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics* 14/3: 275–311.

Biber, Douglas and Frederica Barbieri. 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26/3: 263–286.

Biber, Douglas and Susan Conrad. 1999. Lexical bundles in conversation and academic prose. *Language and Computers* 26: 181–190.

Biber, Douglas, Susan Conrad and Viviana Cortes. 2003. Lexical bundles in speech and writing: An initial taxonomy. In Andrew Wilson, Paul Rayson and Tony McEnery eds. *Corpus Linguistics by the Lune.* Bern: Peter Lang, 71–93.

Biber, Douglas, Susan Conrad and Viviana Cortes. 2004. If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25/3: 371–405.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English.* London: Longman.

Bird, Steven, Ewan Klein and Edward Loper. 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit.* Sebastopol, California: O'Reilly Media Inc.

Bybee, Joan. 2008. Usage-based grammar and second language acquisition. In Peter Robinson and Nick C. Ellis eds. *Handbook of Cognitive Linguistics and Second Language Acquisition.* London: Routledge, 226–246.

Chen, Yu-Hua and Paul Baker. 2010. Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology* 14/2: 30–49.

Cortes, Viviana. 2004. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23/4: 397–423.

Durrant, Philip and Norbert Schmitt. 2009. To what extent do native and non-native writers make use of collocations? *IRAL-International Review of Applied Linguistics in Language Teaching* 47/2: 157–177.

Ellis, Nick C. 2002. Frequency effects in language processing. *Studies in Second Language Acquisition* 24/2: 143–148.

Ellis, Nick C. 2008. The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The Modern Language Journal* 92/2: 232–249.

Esfandiari, Rajab, and Fatima Barbary. 2017. A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes* 29: 21–42.

Gilquin, Gaëtanelle. 2008. Combining contrastive and interlanguage analysis to apprehend transfer: Detection, explanation, evaluation. In Gaëtanelle Gilquin, Szilvia Papp and María Belén Díez-Bedmar eds. *Linking up Contrastive and Learner Corpus Research.* Leiden: Brill, 1–33.

Gilquin, Gaëtanelle and Magali Paquot. 2008. Too chatty: Learner academic writing and register variation. *English Text Construction* 1/1: 41–61.

Gilquin, Gaëtanelle, Sylviane Granger and Magali Paquot. 2007. Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes* 6/4: 319–335.

Granger, Sylviane. 1998. Prefabricated patterns in advanced EFL writing: Collocations and formulae. In Anthony P. Cowie ed. *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press, 145–160.

Granger, Sylviane. 2002. A bird's-eye view of learner corpus research. In Sylviane Granger, Joseph Hung and Stephanie Petch-Tyson eds. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching.* Amsterdam: John Benjamins, 3–33.

Granger, Sylviane and Fanny Meunier. 2008. Phraseology in foreign language learning and teaching. In Fanny Meunier and Sylviane Granger eds. *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins, 15–20.

Granger, Sylviane and Paul Rayson. 1998. Automatic profiling of learner texts. In Sylviane Granger ed. *Learner English on Computer*. London: Routledge, 119–131.

Güngör, Fatih and Hacer Hande Uysal. 2020. Lexical bundle use and crosslinguistic influence in academic texts. *Lingua* 242: 102859. https://doi.org/10.1016/j.lingua.2020.102859

Hyland, Ken. 2008. As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27/1: 4–21.

Hyland, Ken. 2015. *Teaching and Researching Writing*. London: Routledge.

Jaeger, T. Florian and Harry Tily. 2011. Language processing complexity and communicative efficiency. *WIRE: Cognitive Science* 2/3: 323–335.

Jarvis, Scott. 2000. Methodological rigor in the study of transfer: Identifying L1 influence in them interlanguage lexicon. *Language Learning* 50/2: 245–309.

Lu, Xiaofei and Jinlei Deng. 2019. With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students. *Journal of English for Academic Purposes* 39: 21–36.

Nekrasova, Tatiana M. 2009. English L1 and L2 speakers' knowledge of lexical bundles. *Language Learning* 59/3: 647–686.

Odlin, Terrence. 2003. Cross-linguistic influence. In Catherine J. Doughty and Michael H. Long eds. *The Handbook of Second Language Acquisition*. Oxford: Blackwell Publishing, 436–486.

O'Donnell, Matthew Brook, Ute Römer and Nick C. Ellis. 2013. The development of formulaic sequences in first and second language writing. *International Journal of Corpus Linguistics* 18/1: 83–108.

Pan, Fan, Randi Reppen and Douglas Biber. 2016. Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes* 21: 60–71.

Paquot, Magali. 2013. Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics* 18/3: 391–417.

Paquot, Magali. 2014. Cross-linguistic influence and formulaic language: Recurrent word sequences in French learner writing. *EUROSLA Yearbook* 14/1: 240–261.

Paquot, Magali. 2017. L1 frequency in foreign language acquisition: Recurrent word combinations in French and Spanish EFL learner writing. *Second Language Research* 33/1: 13–32.

Pérez-Llantada, Carmen. 2014. Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes* 14: 84–94.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing. R* Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Ren, Junqiang. 2021. Variability and functions of lexical bundles in research articles of applied linguistics and pharmaceutical sciences. *Journal of English for Academic Purposes* 50, 100968. https://doi.org/10.1016/j.jeap.2021.100968

Rica Peromingo, Juan Pedro. 2012. Corpus analysis and phraseology: Transfer of multi-word units. *Linguistics and the Human Sciences* 6/1–3: 321–343.

Römer, Ute. 2009. The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics* 7/1: 140–162.

Salazar, Danica. 2011. *Lexical Bundles in Scientific English: A Corpus-based Study of Native and Non-native Writing*. Barcelona: University of Barcelona dissertation.

Salazar, Danica. 2014. *Lexical Bundles in Native and Non-native Scientific Writing: Applying a Corpus-based Study to Language Teaching*. Amsterdam: John Benjamins.

Selinker, Larry. 1966. *A Psycholinguistic Study of Language Transfer*. Washington: Georgetown University dissertation.

Simpson-Vlach, Rita and Nick C. Ellis. 2010. An academic formulas list: New methods in phraseology research. *Applied Linguistics* 31/4: 487–512.

Staples, Shelley, Jesse Egbert, Douglas Biber and Alyson McClair. 2013. Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes* 12/3: 214–225.

Swales, John M. 2005. Attended and unattended "this" in academic writing: A long and unfinished story. *ESP Malaysia* 11/1: 1–15.

Swales, John M. and Christine B. Feak. 2004. *Academic Writing for Graduate Students: Essential Tasks and Skills, Vol. 1*. Michigan: University of Michigan Press.

Wulff, Stefanie and Ute Römer. 2009. Becoming a proficient academic writer: Shifting lexical preferences in the use of the progressive. *Corpora* 4/2: 115–133.

Corresponding author
Maria Kostromitina
Department of English
Northern Arizona University
705 S Beaver St.,
Flagstaff, AZ, 86011, USA
e-mail: mk782@nau.edu

APPENDICES

Appendix 1: Complete lexical bundle list.

PSY-ENG1

Rank	Frequency	Freq. per mil.	Range	Bundle
1	122	294	9	*the online supplemental materials*
2	79	190	7	*in the online supplemental*
3	56	135	26	*in the context of*
4	49	118	26	*it is important to*
5	33	79	20	*are more likely to*
6	28	67	17	*the extent to which*
7	25	60	20	*as well as the*
8	25	60	13	*more likely to be*
9	24	58	8	*online supplemental materials for*
10	23	55	15	*a wide range of*
11	21	48	10	*one of the most*
12	21	50	6	*see the online supplemental*
13	18	43	8	*the degree to which*
14	17	41	10	*a meta analysis of*
15	17	41	9	*in the form of*
16	16	38	9	*as a function of*
17	16	38	9	*as part of a*
18	16	38	8	*at the end of*
19	16	38	8	*in the development of*
20	16	38	9	*it is possible that*
21	16	38	6	*with respect to the*
22	15	36	8	*has been shown to*
23	15	36	11	*have been shown to*
24	15	36	8	*health and well being*
25	15	36	9	*(is) important to note that +*
26	15	36	9	*were more likely to*
27	14	33	7	*can be used to*
28	14	33	11	*in addition to the*
29	14	33	6	*(the) science and practice of +*
30	14	33	7	*the nature of the*
31	14	33	8	*the ways in which*
32	13	31	11	*as a result of*
33	13	31	8	*national institutes of health*
34	13	31	11	*of this article is*
35	13	31	6	*the context of the*
36	13	31	11	*to the extent that*
37	12	29	10	*in a sample of*
38	12	29	10	*in a way that*
39	12	29	8	*in the general population*
40	12	29	9	*research is needed to*
41	12	29	7	*we were able to*
42	11	25	7	*across the life span*

43	11	25	9	*been shown to be*
44	11	25	9	*has the potential to*
45	11	25	9	*in terms of the*
46	11	25	8	*in the case of*
47	11	25	9	*it may also be*
48	11	25	8	*over the past years*
49	10	24	9	*a wide variety of*
50	10	24	7	*at the university of*
51	10	24	8	*in the absence of*
52	10	24	9	*is one of the*
53	10	24	8	*it is possible to*
54	10	24	8	*physical and mental health*
55	10	24	8	*(the) purpose of this article (is to) +*
56	9	21	7	*a risk factor for*
57	9	21	6	*and physical well being*
58	9	21	8	*and the development of*
59	9	21	6	*in the face of*
60	9	21	7	*in this article we*
61	9	21	7	*is likely to be*
62	9	21	8	*it is clear that*
63	9	21	6	*it should be noted (that) +*
64	9	21	6	*the importance of the*
65	9	21	6	*the magnitude of the*
66	8	20	7	*a wide array of*
67	8	20	6	*as part of the*
68	8	20	7	*at the same time*
69	8	20	6	*at the time of*
70	8	20	8	*for the development of*
71	8	20	6	*has focused on the*
72	8	20	6	*in light of the*
73	8	20	6	*it may be that*
74	8	20	6	*of health and human*
75	8	20	7	*over a year period*
76	8	20	7	*research has shown that*
77	8	20	6	*the full range of*
78	8	20	6	*the national institutes of*
79	8	20	8	*to be associated with*
80	8	20	8	*was associated with a*
81	8	20	7	*with a focus on*
82	8	20	7	*within the context of*

PSY-ENG2

Rank	Frequency	Freq. per mil.	Range	Bundle
1	96	303	41	*at the same time*
2	72	224	18	*a high level of*
3	57	180	34	*the results of the*
4	55	170	27	*as well as the*
5	54	168	20	*in the case of*
6	48	151	26	*on the basis of*
7	46	145	30	*on the other hand*
8	45	142	26	*in the context of*
9	45	142	21	*in the process of*
10	36	113	22	*is one of the*
11	35	110	20	*it is necessary to*
12	35	110	6	*with a high level*
13	34	107	6	*russian version of the*
14	32	101	16	*the relationship between the*
15	32	101	18	*with the help of*
16	31	97	20	*one of the most*
17	30	94	25	*on the one hand*
18	29	91	18	*(it) should be noted that +*
19	28	88	9	*a higher level of*
20	27	85	12	*for the development of*
21	26	82	22	*is based on the*
22	26	82	13	*the end of the*
23	25	78	17	*the fact that the*
24	24	75	15	*in the course of*
25	24	75	16	*it is important to*
26	23	72	14	*as a result of*
27	23	72	18	*in the form of*
28	23	72	15	*the analysis of the*
29	22	69	11	*in the field of*
30	22	69	15	*it is possible to*
31	22	69	12	*on the level of*
32	21	66	13	*it was found that*
33	20	63	10	*a low level of*
34	20	63	14	*at the end of*
35	20	63	9	*in the structure of (the) +*
36	20	63	12	*that there is a*
37	19	60	13	*turned out to be*
38	18	56	11	*are presented in table*
39	18	56	11	*the basis of the*
40	18	56	13	*the level of the*
41	18	56	14	*the same time the*
42	18	56	14	*to the fact that*
43	17	53	12	*a number of studies*
44	17	53	14	*in accordance with the*

45	17	53	8	*in the group of*
46	17	53	8	*level of development of*
47	17	53	11	*the first stage of*
48	16	50	12	*be explained by the*
49	16	50	12	*in this case the*
50	16	50	7	*the case of the*
51	16	50	12	*the development of the*
52	16	50	11	*to the study of*
53	15	47	9	*of the relationship between*
54	15	47	9	*studies have shown that*
55	15	47	6	*the content of the*
56	15	47	10	*the study of the*
57	15	47	10	*the total number of*
58	15	47	7	*with different levels of*
59	14	44	11	*can be explained by*
60	14	44	11	*in the study of*
61	14	44	7	*of the level of*
62	14	44	11	*of this study is*
63	14	44	11	*the beginning of the*
64	14	44	7	*the dynamics of the*
65	14	44	9	*to the development of*
66	13	41	8	*and the level of*
67	13	41	8	*as one of the*
68	13	41	9	*at the level of*
69	13	41	12	*by the fact that*
70	13	41	11	*is considered to be*
71	13	41	8	*of the development of*
72	13	41	9	*the development of a*
73	13	41	9	*the purpose of this*
74	13	41	9	*to the conclusion that*
75	12	38	9	*as well as to*
76	12	38	8	*(it) can be assumed that +*
77	12	38	10	*in contrast to the*
78	12	38	8	*in other words the*
79	12	38	7	*in the learning process*
80	12	38	9	*in the number of*
81	12	38	6	*in the present study*
82	12	38	10	*it is possible that*
83	12	38	9	*the context of the*
84	12	38	7	*the same time they*
85	12	38	9	*the value of the*
86	11	35	7	*an increase in the*
87	11	35	7	*and the development of*
88	11	35	10	*as well as in*
89	11	35	9	*for each of the*
90	11	35	11	*in a number of*

91	11	35	8	*in the development of*
92	11	35	10	*makes it possible to*
93	11	35	9	*of this study was*
94	11	35	7	*one of the first*
95	11	35	10	*results of the study*
96	11	35	10	*the rest of the*
97	11	35	11	*the role of the*
98	11	35	9	*took part in the*
99	11	35	7	*with respect to the*
100	10	32	6	*as in the case*
101	10	32	9	*due to the fact*
102	10	32	8	*in the fact that*
103	10	32	8	*is consistent with the*
104	10	32	9	*is determined by the*
105	10	32	9	*is related to the*
106	10	32	7	*it was shown that*
107	10	32	6	*of the dynamics of*
108	10	32	9	*of the most important*
109	10	32	9	*point of view of*
110	10	32	6	*the concept of the*
111	10	32	8	*the formation of the*
112	10	32	7	*the influence of the*
113	10	32	6	*the meaning of the*
114	10	32	8	*was found that the*
115	10	32	7	*we can assume that*
116	10	32	8	*we can conclude that*
117	10	32	8	*we can say that*
118	10	32	7	*with the results of*
119	10	32	7	*with the use of*
120	9	28	7	*an important role in*
121	9	28	6	*and the degree of*
122	9	28	8	*as a basis for*
123	9	28	8	*as well as a*
124	9	28	9	*at the beginning of*
125	9	28	6	*at the time of*
126	9	28	7	*can be found in*
127	9	28	7	*from the perspective of*
128	9	28	8	*in addition to the*
129	9	28	7	*in terms of the*
130	9	28	8	*in the works of*
131	9	28	7	*of the fact that*
132	9	28	8	*take into account the*
133	9	28	8	*the characteristics of the*
134	9	28	7	*the differences between the*
135	9	28	7	*the nature of the*
136	9	28	9	*the result of the*

137	9	28	9	*the results of our*
138	9	28	6	*the same time it*
139	9	28	8	*the use of the*
140	9	28	8	*the validity of the*
141	9	28	7	*this study was to*
142	9	28	6	*to be the most*
143	9	28	8	*was based on the*
144	9	28	7	*which is based on*
145	9	28	6	*within the framework of*
146	8	25	6	*a negative impact on*
147	8	25	7	*a result of the*
148	8	25	8	*and at the same*
149	8	25	8	*and on the other*
150	8	25	8	*and the ability to*
151	8	25	8	*as well as their*
152	8	25	7	*be noted that the*
153	8	25	7	*can serve as a*
154	8	25	6	*did not differ from*
155	8	25	7	*explained by the fact*
156	8	25	8	*in front of the*
157	8	25	7	*in our case the*
158	8	25	6	*in our opinion the*
159	8	25	6	*in our research we*
160	8	25	6	*in the current study*
161	8	25	6	*in the educational process*
162	8	25	7	*in which a person*
163	8	25	6	*of the ability to*
164	8	25	6	*of the study we*
165	8	25	6	*on the development of*
166	8	25	6	*the aim of the*
167	8	25	6	*the conclusion that the*
168	8	25	8	*the one hand the*
169	8	25	8	*the quality of the*
170	8	25	8	*the results of this*
171	8	25	7	*the second stage of*
172	8	25	8	*there were no significant*
173	8	25	7	*this point of view*
174	8	25	8	*to take into account*
175	8	25	8	*under the influence of*
176	8	25	7	*us to conclude that*
177	7	22	6	*a great number of*
178	7	22	6	*an analysis of the*
179	7	22	6	*and as a result*
180	7	22	6	*and the number of*
181	7	22	6	*as the result of*
182	7	22	7	*can conclude that the*

183	7	22	6	*considered to be an*
184	7	22	6	*correlation analysis of the*
185	7	22	6	*for the first time*
186	7	22	6	*for the study of*
187	7	22	7	*in order to achieve*
188	7	22	6	*in our study we*
189	7	22	6	*in the same way*
190	7	22	6	*in this study the*
191	7	22	6	*is due to the*
192	7	22	6	*is understood as a*
193	7	22	7	*it turned out that*
194	7	22	7	*of the results of*
195	7	22	7	*of the study was*
196	7	22	7	*on the results of*
197	7	22	7	*one of the main*
198	7	22	6	*significant differences in the*
199	7	22	6	*the differences in the*
200	7	22	6	*the idea of the*
201	7	22	6	*the importance of the*
202	7	22	6	*the other hand the*
203	7	22	6	*the point of view*
204	7	22	6	*the research was conducted*
205	7	22	6	*to the analysis of*
206	6	20	6	*a high degree of*
207	6	20	6	*and the results of*
208	6	20	6	*at the age of*
209	6	20	6	*considered to be the*
210	6	20	6	*during the process of*
211	6	20	6	*for a long time*
212	6	20	6	*in line with the*
213	6	20	6	*in the formation of*
214	6	20	6	*is associated with the*
215	6	20	6	*make it possible to*
216	6	20	6	*not only in the*
217	6	20	6	*of the study is*
218	6	20	6	*one of the key*
219	6	20	6	*results of this study*
220	6	20	6	*that the role of*
221	6	20	6	*the form of a*
222	6	20	6	*the one hand and*
223	6	20	6	*the reliability of the*
224	6	20	6	*the study was to*
225	6	20	6	*to a lesser extent*
226	6	20	6	*to the theory of*
227	6	20	6	*turned out that the*

PSY-RUS1

Rank	Freq	Freq. per mil.	Range	Bundle	Translation
1	232	758	58	*о том что*	*about this [the fact] that*
2	103	336	30	*по сравнению с*	*in comparison with*
3	86	281	37	*в том что*	*in this [the fact] that*
4	81	248	34	*и т п*	*and similar*
5	79	242	39	*в том числе*	*including*
6	74	238	33	*в связи с*	*in connection to*
7	74	238	30	*и т д*	*and so on*
8	66	215	29	*в зависимости от*	*in dependence with/ depending on*
9	66	215	36	*в отличие от*	*in contrast with*
10	60	196	29	*в соответствии с*	*in agreement with/in contrast with*
11	60	196	36	*на то что*	*to this [the fact] that*
12	58	189	29	*вместе с тем*	*at the same time/together with this*
13	55	179	28	*с точки зрения*	*from the point of view*
14	51	166	27	*по отношению к*	*in relation to*
15	51	166	21	*тех или иных*	*these or others*
16	48	156	22	*в первую очередь*	*in first turn/ firstly*
17	48	156	23	*в то же*	*at the same*
18	47	153	25	*с одной стороны*	*from the one side/ on the one hand*
19	46	150	22	*на наш взгляд*	*in our view*
20	46	150	23	*то же время*	*at the same time*
21	46	150	23	*(в) то же время +*	*at the same time*
22	44	143	9	*в социальных сетях*	*in social networks*
23	44	143	26	*тот факт что*	*the fact that*
24	40	130	24	*следует отметить что*	*needed to point out that*
25	39	127	27	*в то время*	*at the time/ while*
26	38	124	20	*тем не менее*	*nevertheless*
27	37	120	15	*было показано что*	*was shown that*
28	37	120	28	*в настоящее время*	*at present time/ currently*
29	36	117	17	*в большей степени*	*to a greater extent/degree*
30	36	117	17	*в свою очередь*	*in its turn*
31	35	114	20	*связи с этим*	*connection with this*
32	34	111	23	*то время как*	*at the time when*
33	35	111	20	*(в) связи с этим+*	*concerning that/ in connection to*
34	33	108	20	*в этом случае*	*in this case*
35	34	107	23	*в то время как*	*while/ at the same time as*
36	32	105	21	*с другой стороны*	*from the other side/ on the other hand*
37	32	105	9	*там же с*	*also there with*
38	32	104	16	*вывод о том что +*	*conclusion that / conclusion about the fact that*

39	28	91	23	в данном случае	in this case
40	28	91	19	можно предположить что	can assume that [it can be assumed that]
41	30	90	30	описание хода исследования	study process description
42	27	88	18	речь идет о	talk is about/ this is about
43	27	88	18	с тем что	with this [idea] that
44	27	88	14	сделать вывод о	make a conclusion about
45	25	85	10	как видно из	as seen from
46	25	85	18	так и в	also in/ as well as in
47	24	78	16	в возрасте от	in the age from
48	24	78	12	в ответ на	in response to
49	24	78	6	в реальной жизни	in real life
50	24	78	17	той или иной	this or that
51	23	75	6	на уровне тенденции	on the tendency level
52	22	75	16	а так же	as well as
53	22	75	11	в подростковом возрасте	in adolescent age
54	22	75	15	до сих пор	until now
55	22	75	18	не только	not only
56	21	68	16	включает в себя	includes
57	21	68	8	детей и подростков	children and adolescents
58	21	68	17	для того чтобы	in order to
59	21	68	10	на самом деле	in reality
60	21	68	13	при этом в	at the same time in/ with this in
61	21	68	11	состоит в том	consists of
62	20	65	14	на вопрос о	to the question of
63	20	65	13	так же как	as well as
64	20	65	16	таким образом в	thus/therefore
65	19	62	17	вопрос о том	quiestion about that [the fact that]
66	19	62	12	по нашему мнению	in our opinion
67	18	59	11	в данной работе	in this work
68	18	59	12	на этом этапе	at this stage
69	18	59	11	не может быть	cannot be
70	18	59	16	том что в	this [this fact] that in
71	18	58	9	сделать вывод о том	conclude that/ make a conclusion that
72	17	55	10	вне зависимости от	independent of
73	17	55	14	не только в	not only in
74	17	55	13	но при этом	but at the same time
75	17	55	10	те или иные	these or others
76	17	55	15	том числе и	including and
77	17	55	15	в том числе и	also including
78	17	55	10	состоит в том что	consists of this [the fact] that
79	16	52	10	в нашем исследовании	in our study
80	16	52	7	друг от друга	from each other

81	16	52	12	*и др в*	*and others in*
82	16	52	15	*несмотря на то (что) +*	*despite the fact that*
83	16	52	13	*о том как*	*about how*
84	16	52	15	*несмотря на то*	*despite the fact that*
85	15	49	11	*исследовании приняли участие*	*took part in the study*
86	15	49	10	*можно рассматривать как*	*can be viewed as*
87	15	49	12	*отметить что в*	*note that in*
88	15	49	11	*является одним из*	*is one of the*
89	14	46	10	*так же как и +*	*as well as*
90	14	45	8	*более высокий уровень*	*a higher level*
91	14	45	12	*друг с другом*	*with each other*
92	14	45	9	*зависимости от того*	*depending on*
93	14	45	11	*к тому что*	*to this [the fact] that*
94	14	45	10	*как и в*	*as in/like in*
95	14	45	12	*на первый план*	*in the foreground*
96	14	45	10	*на сегодняшний день*	*nowadays/ to date*
97	14	45	8	*с другими людьми*	*with other people*
98	14	45	10	*так и на*	*as well as on*
99	14	45	11	*того или иного*	*that or the other [gen]*
100	14	45	11	*является одной из*	*is one of*
101	13	43	9	*в исследовании приняли участие*	*took part in the study*
102	13	43	10	*заключается в том что*	*consists in this [the fact] that*
103	13	42	9	*а с другой*	*and on the other*
104	13	42	9	*в исследовании приняли*	*in the study took*
105	13	42	10	*заключается в том*	*can be summarized in*
106	13	42	9	*и тем самым*	*and with that*
107	13	42	10	*по всей видимости*	*evidently/apparently*
108	13	42	11	*после того как*	*after this [the fact] that*
109	12	40	9	*в той или иной*	*in one or another*
110	12	40	6	*говорить о том что*	*talk about*
111	12	40	8	*с нашей точки зрения +*	*from our point of view*
112	12	40	8	*свидетельствует о том что*	*indicates that*
113	12	40	9	*свидетельствуют о том что +*	*indicate that*
114	12	39	6	*было установлено что*	*was established that*
115	12	39	8	*в данном исследовании*	*in the present study*
116	12	39	9	*в той или*	*in this or*
117	12	39	6	*говорить о том*	*talk about*
118	12	39	10	*можно сделать вывод*	*can be concluded*
119	12	39	6	*особый интерес представляет*	*presents special interest*
120	12	39	10	*с высоким уровнем*	*with a high level of*

					provides evidence to [sing]/ indicates that/ indicates the
121	12	39	8	свидетельствует о том	fact that
122	12	39	10	так и для	as well as for/to
123	12	39	10	таким образом	therefore
124	12	39	12	хода исследования в	study process in
125	12	39	6	юношей и девушек	young men and women
126	10	37	9	в том что в	in that/ in the [fact] that
127	10	37	8	можно сделать вывод о	we can conclude/ can be concluded ...about
128	10	37	7	обращает на себя внимание	noteworthy/draws attention upon itself
129	11	35	7	а также на	as well as on
130	11	35	8	а также с	as well as with
131	11	35	8	более или менее	more or less
132	11	35	7	в меньшей степени	to a lesser degree/ extent
133	11	35	11	в нашей работе	in our work
134	11	35	9	в работе с	in the work with
135	11	35	9	исследования показали что	of the study showed that
136	11	35	8	можно говорить о	can be talked about
137	11	35	7	на себя внимание	attention on itself
138	11	35	8	необходимо отметить что	it is necessary to note that
139	11	35	6	по всей выборке	throughout the sample
140	11	35	10	по крайней мере	at least
141	11	35	7	представлены в таблице	are presented in table
142	11	35	10	то что в	this [the fact] that in
143	9	33	6	в том числе в	including in
144	9	33	6	вне зависимости от того	no matter/ independent of [the fact] that
145	9	33	7	как видно из таблицы	as can be seen from table
146	10	32	8	а также в	as well as in
147	10	32	9	в данной статье	in the present article
148	10	32	6	в обеих группах	in both groups
149	10	32	6	в отечественной психологии	in the fatherland psychology
150	10	32	9	в последнее время	recently
151	10	32	6	в сочетании с	together with / in conjunction with
152	10	32	7	в том случае	in the case of
153	10	32	9	в частности в	in particular in
154	10	32	9	важно отметить что	it is important to note that
155	10	32	8	видно из таблицы	seen from table
156	10	32	8	и в целом	and in general
157	10	32	6	на этой стадии	at this stage
158	10	32	7	обращает на себя	draws upon itself
159	10	32	8	по их мнению	in their opinion
160	9	29	7	в конечном счете	in the end [as a result]

161	9	29	7	в некоторых случаях	in some cases
162	9	29	7	в полной мере	fully/ in full capacity
163	9	29	6	в этой области	in this area
164	9	29	6	и уверенность в	and confidence in
165	9	29	7	из того что	from [the fact] that
166	9	29	8	кроме того в	Besides, in
167	9	29	6	могут быть связаны	may be related
168	9	29	8	может привести к	can lead to
169	9	29	6	при этом не	while not
170	9	29	8	при этом они	while they
171	9	29	9	с опорой на	based on
172	9	29	9	с помощью	via/with the help of
173	9	29	8	с таким образом	with that way
174	9	29	7	так или иначе	anyway/ this or that way
175	9	29	9	таким образом можно	this way you can/we can/ it is possible
176	9	29	6	том числе в	including in
177	9	29	8	том что они	that they
178	9	29	6	человека и его	man and his
179	9	29	6	что по мере	that as
180	9	29	9	это может быть	it could be
181	9	29	7	это означает что	it means that
182	8	29	7	и т д в	and so on in
183	8	29	7	одни и те же +	same
184	8	29	7	указывает на то что	indicates that / [the fact] that/ points to the fact that
185	8	29	7	что в свою очередь	which in turn
186	8	26	7	было выявлено что	it was revealed that
187	8	26	6	в какой то	at some
188	8	26	6	в рамках данного	as part of this/in the frame o this
189	8	26	6	в ряде случаев	in some cases
190	8	26	7	в том чтобы	in that, to
191	8	26	6	в целом по	on the whole
192	8	26	8	важную роль в	important role in
193	8	26	7	друг к другу	to each other
194	8	26	8	и при этом	and wherein
195	8	26	6	и таким образом	and thus
196	8	26	7	и те же	the same
197	8	26	8	и то что	and [the fact] that
198	8	26	8	может быть связано	may be related
199	8	26	6	можно сказать что	we can say that
200	8	26	7	на то чтобы	in order to
201	8	26	6	не могут быть	can not be
202	8	26	7	не только на	not only on
203	8	26	8	но и в	but also in

204	8	26	6	по мере увеличения	as you increase/as we increase
205	8	26	7	с ним в	with him in
206	8	26	7	становится все более	getting more/becoming increasingly
207	8	26	7	так и не	as well as not
208	8	26	7	так и с	as well as with
209	8	26	7	указывает на то	indicates that
210	8	26	7	что в свою	which in its
211	8	26	7	что может быть	what could be
212	8	26	7	что он не	that he not
213	7	25	6	в том случае если	in case if
214	7	25	6	вместе с тем в	at the same time in
215	7	25	6	внимание на то (что)+	attention to [the fact] that about that in/ about [the fact] that
216	7	25	6	о том что в	
217	7	23	7	а также о	as well as about
218	7	23	6	в значительной степени	to a large extent
219	7	23	6	в конце х	at the end of x
220	7	23	6	в одном из	in one of
221	7	23	7	в последние годы	in recent years
222	7	23	7	в том числе	including
223	7	23	7	в целом и	in general and
224	7	23	6	внимание на то	attention to [the fact]
225	7	23	7	время от времени	occasionally/ from time to time
226	7	23	6	исследование показало что	the study showed that
227	7	23	7	их связи с	their relationship with
228	7	23	6	мы предположили что	we assumed that
229	7	23	6	на этот вопрос	to this question
230	7	23	6	предположить что в	suggest that in
231	7	23	6	при этом у	at the same time in
232	7	23	7	равно как и	as well as
233	7	23	6	с тем в	with that in
234	7	23	7	связано с тем	due to the/ connected to the
235	7	23	6	сделать следующие выводы	draw the following conclusions
236	7	23	6	том случае если	in case if
237	7	23	7	это связано с	it's connected with
238	6	20	6	р при этом	p in this case
239	6	20	6	а не на	and not on
240	6	20	6	а также их	as well as their
241	6	20	6	в исследовании были	in the study were
242	6	20	6	в которой он	in which he
243	6	20	6	в последние десятилетия	in recent decades
244	6	20	6	в целом в	generally in
245	6	20	6	до того как	before

246	6	20	6	*и в отношении исследования в*	*and regarding/and in relation to*
247	6	20	6	*исследовании*	*research in research*
248	6	20	6	*исследования в статье*	*research in the article*
249	6	20	6	*к тому же*	*in addition*
250	6	20	6	*как раз и*	*just and*
251	6	20	6	*можно выделить несколько*	*there are several/it is possible to single out (highlight) several*
252	6	20	6	*мы предполагаем что*	*we assume that*
253	6	20	6	*мы считаем что*	*we believe that*
254	6	20	6	*на первом этапе*	*at the first stage*
255	6	20	6	*на этой основе*	*on this basis*
256	6	20	6	*не только для*	*not only for*
257	6	20	6	*но и на*	*but also on*
258	6	20	6	*отметить что на*	*note that on*
259	6	20	6	*по его мнению*	*in his opinion*
260	6	20	6	*тем или иным*	*one way or another*
261	6	20	6	*том что он*	*that he [the fact] that he*
262	6	20	6	*человек в возрасте*	*elderly person*
263	6	20	6	*это проявляется в*	*it manifests itself in*
264	6	20	6	*связано с тем что*	*due to [the fact] that/connected to [the fact] that*

Appendix 2: Functional classification of the first 50 frequent bundles

---

Research-oriented

Description	PSY-ENG1 *the online supplemental materials, online supplemental materials for, in the development of, the nature of the, in a way that, the ways in which*
	PSY-ENG2 *the relationship between the, is based on the, for the development of in the form of, in the field of, on the level of, turned out to be, the basis of the, the level of the, level of development of, be explained by the*
	PSY-RUS1 *в возрасте от / in the age from, той или иной / that or other*
Location	PSY-ENG1 *in the online supplemental, at the end of, over the past years, at the university of*
	PSY-ENG2 *the end of the, in the course of, at the end of, in the structure of (the)+, that there is a, in the group of*
	PSY-RUS1 *в настоящее время / at present time, там же с / also there with*
Quantification	PSY-ENG1 *the extent to which, a wide range of, one of the most, the degree to which, to the extent that, a wide variety of*
	PSY-ENG2 *a high level of, is one of the, with a high level, one of the most, a higher level of, a low level of, a number of studies*
	PSY-RUS1 *в большей степени / to a greater extent*
Procedure	PSY-ENG1 *a meta analysis of, as part of a, in the general population, we were able to*
	PSY-ENG2 *the results of the, in the process of, with the help of, the analysis of the, the first stage of*
	PSY-RUS1 *в ответ на / in response to, в то время / at the time*
Topic	PSY-ENG1 *health and wellbeing of, (the) science and practice of +, national institutes of health, across the lifespan of*
	PSY-ENG2 *russian version of the*
	PSY-RUS1 *в социальных сетях / in social networks, в реальной жизни / in real life*

Text-oriented

Framing signals	PSY-ENG1 *in the context of, in the form of, as a function of, with respect to the, the context of the, in the case of*

PSY-ENG2
in the case of, on the basis of, in the context of, in this case the, the case of the

PSY-RUS1
в том числе / including, в зависимости от / depending on, по отношению к / in relation to, речь идет о / the talk is about

Transition signals	PSY-ENG1: *as well as the, in addition to the, in terms of the*
	PSY-ENG2: *at the same time, as well as the, on the other hand, on the one hand, the same time the, in accordance with the*
	PSY-RUS1: *в то время как / however, о том что / about this [the fact] that, по сравнению с / in, comparison to, в том что / in [the fact] that, и m n / and so on, в связи с / in connection to, и m д / and so forth, в отличие от / in contrast with, в соответствии с / in agreement with, на то что / to [the fact] that, вместе с тем / together with this, с точки зрения / from the point of view, в первую очередь / firstly, с одной стороны / on the one hand, в то же время / at the same time, тем не менее / nevertheless, в свою очередь / in its turn, (в) связи с этим + / in connection to this, с другой стороны / on the other hand, с тем что (more specifically), так и в / so in*
Structuring signals	PSY-ENG1: *see the online supplemental, of this article is, in a sample of*
	PSY-ENG2: *are presented in table*
	PSY-RUS1: *описание хода исследования / study process description, как видно из / as seen from*
Resultative signals	PSY-ENG1: *has been shown to, have been shown to, as a result of*
	PSY-ENG2: *as a result of, it was found that*
	PSY-RUS1: *вывод о том что + / conclusion about [the fact] that, сделать вывод о / make a conclusion about*

Stance-oriented

Engagement features	PSY-ENG1: *it is important to, (is) important to note that +*
	PSY-ENG2: *it is necessary to, (it) should be noted that +, it is important to*
	PSY-RUS1: *можно предположить что / it can be assumed that следует отметить что / should be pointed out that*
Stance features	PSY-ENG1: *are more likely to, more likely to be, it is possible that, were more likely to, can be used to, research is needed to, been shown to be, has the potential to, it may also be*
	PSY-ENG2: *the fact that the, it is possible to, to the fact that*
	PSY-RUS1: *тот факт что / the fact that, на наш взгляд / in our view*